Version of Amended Specification and Claims
With Markings To Show Changes Made
Case No. 13451US02
Serial No. 09/775,476
Filed Feb. 2, 2001

In the specification:

Kindly delete the paragraph beginning "Figure 2" at page 10, line 28 and ending ""STABLE state" at page 11, line 18, and replace it with the following paragraph:

Figure 2 illustrates one implementation of memory cell 103 in Figure 1, in the form of six-transistor CMOS cell 200. Transistor cell 200 is one type of transistor which also may be used in embodiments of the present invention. SRAM cell 200 can be in one of three possible states: (1) the STABLE state, in which cell 200 holds a signal value corresponding to a logic "1" or logic "0"; (2) a READ operation state; or (3) a WRITE operation state. In the STABLE state, memory cell 200 is effectively disconnected from the memory core (e.g., core 102 in FIG. 1). Bitlines BIT 202[,] and $\overline{BIT}$ 204 are precharged HIGH (logic "1") before any operation (READ or WRITE) can take place. Row select transistors 206, 208 are turned off during precharge. Precharge power is supplied by precharge cells (not shown) coupled with the bitlines 202, 204, similar to precharge cells 120 in Figure 1. A READ operation is initiated by performing a PRECHARGE cycle, precharging bitlines 202, 204 to logic HIGH, and activating word line 205 using row select transistors 206, 208. One of the bitlines 202, 204 discharges through bit cell 200, and a differential voltage is setup between the bitlines 202, 204. This voltage is sensed and amplified to logic levels. A WRITE operation to cell 200 is carried out after another PRECHARGE cycle, by driving bitlines 202, 204 to the required state, and activating word line WORD 205. CMOS is a desirable technology because

the supply current drawn by such an SRAM cell typically is limited to the leakage current of transistors 201a-d while in the STABLE state.

Kindly delete the paragraph beginning "Hierarchical memory" at page 30, line 19 and ending "can be realized" at page 31, line 16, and replace it with the following paragraph:

Hierarchical memory structures also can employ local word line decoding, as illustrated in memory device 740. In FIG. 7, memory device 740 is the uppermost tier (L_2) in the hierarchical memory structure, thus incoming global word line signal 746 is received from global word line drivers (not shown) such as global row address decoders 110 in FIG. 1. In certain preferred embodiments of the present invention, predecoding is employed to effect rapid access to desired word lines, although predecoding is not required, and may not be desired, at every tier in a particular implementation. Signal M46 is received by upper tier predecoder 747, predecoded and supplied to upper tier (L_2) global word line decoders, such as global word line decoder 748. Decoder [M48]748 is coupled with local word line decoder 749 by way of upper tier global word line 750, and selectively activates upper tier local word line decoder 749. Activated L_2 local decoder [M49]749, in turn, activates L_2 local word line 751, which propagates selected word line signal 726 to intermediate tier (L_1) predecoder 727. Predecoder 727 decodes and activates the appropriate intermediate tier (L_1) global word line decoder, such as global word line decoder 728. Decoder 728 is coupled with, and selectively activates, tier L_1 local word line decoder 729 by way of tier (L_1) global word line 730. Activated L_1 local decoder 729, in turn, propagates a selected word line signal 706 to fundamental tier (L_0) predecoder 707, which decodes and activates the appropriate tier

L_0 global word line decoder, such as global word line decoder 708. Activated L_0 local decoder 709, in turn, activates L_0 local word line 711, and selects memory cell 701 for access. In view of the foregoing discussion of hierarchical word line decoding, a skilled artisan would realize that "local word line decoding" occurs at each tier L_0, L_1, and L_2 in the exemplary hierarchy, and is desirable because a substantial reduction in the time and power needed to access selected memory cells can be realized.

Kindly delete the paragraph beginning "In general," at page 34, line 2 and ending "gain adaptively" at page 34, line 33, and replace it with the following paragraph:

In general, single-ended sense amplifiers are useful to save metal space, however, existing designs tend not to be robust due to their susceptibility to power supply and ground noise. In yet another aspect of the present invention, FIG. 10 illustrates a single-ended sense amplifier 1000, preferably with a sample-and-hold reference. Amplifier 1000 can be useful, for example, as a global sense amplifier, sensing input data. At the beginning of an operation, DataIn 1004 is sampled, preferably just before the measurement begins. Therefore, supply, ground, or other noise will affect the reference voltage of sense amplifier 1000 generally in the same way noise affects node to be measured, tending to increase the noise immunity of the sense amplifier 1000. Both inputs 1010, 1011 of differential amplifier 1012 are at the voltage level of DataIn 1004 when the activate signal (GWSELH) 1014 is logic LOW (i.e., at zero potential). At a preselected interval before the measurement begins, but before DataIn [1013]1004 begins to change, activate signal (GWSELH) 1014 is asserted to logic HIGH, thereby isolating the input node 1002 of the transistor [M162]1008. The DataIn voltage existing just before the measurement is taken is sampled and held as a

reference, thereby making the circuit substantially independent of ground or supply voltage references. Transistors [M190] 1025 and [M187] 1026 can add capacitance to the node 1021 where the reference voltage is stored. Transistor [M190] 1025 also can be used as a pump capacitance to compensate for the voltage decrease at the reference node 1021 when the activate signal becomes HIGH and pulls the source 1002 of [M162] 1008 to a lower voltage. Feedback 1030 from output data Data_toLSA 1035, being transmitted to a local sense amplifier (not shown), is coupled with the source/drain of transistor [M187] 1026, actively adjusting the reference voltage at node 1021 by capacitive coupling, thereby adjusting the amplifier gain adaptively.

Kindly delete the paragraph beginning "In yet another" at page 35, line 3 and ending "bitlines 1120, 1121" at page 35, line 33, and replace it with the following paragraph:

In yet another aspect of the present invention, a latch-type sense amplifier 1100 with dynamic offset cancellation is provided. Sense amplifier 1100 also may be useful as a global sense amplifier, and is suited for use in conjunction with hierarchical bitline sensing. Typically, the sensitivity of differential sense amplifiers can be limited by the offsets caused by inherent process variations for devices ("device matching"), and dynamic offsets that may develop on the input lines during high-speed operation. Decreasing the amplifier offset usually results in a corresponding decrease in the minimum bitline swing required for reliable operation. Smaller bitline swings can lead to faster, lower power memory operation. With amplifier 1100, the offset on bitlines can be canceled by the triple PMOS precharge-and-balance transistors [M3] 1101, [M4] 1102, [M5] 1103, which arrangement is known to those skilled in the art. However, despite

precharge-and-balance transistors 1101-1103, an additional offset at the inputs of the latch may exist. By employing balancing PMOS transistors [(M14)] 1110, any offset that may be present at the input of the latch-type differential sense amplifier can be substantially equalized. Sense amplifier 1100 demonstrates a charge-sharing limited swing driver 1115. Global bitlines 1150, 1151 are disconnected from sense amplifier 1100 when sense amplifier 1100 is not being used, i.e., in a tri-state condition. Sense amplifier 1100 can be in a precharged state if both input/output nodes are logic HIGH, i.e., if both of the PMOS drivers, [M38] 1130 and [M29] 1131 are off (inputs at logic HIGH). A large capacitor, [C_0] 1135, in sense amplifier 1100 can be kept substantially at zero volts by two series NMOS transistors, [M37] 1140 and [M40] 1141. The size of capacitor 1135 can be determined by the amount of voltage swing typically needed on global bitlines 1120, 1121.

Kindly delete the paragraph beginning "When sense amplifier" at page 35, line 34 and ending "input and output" at page 36, line 10, and replace it with the following paragraph:

When sense amplifier 1100 is activated, and bitlines 1150, 1151 are logic HIGH, PMOS transistor [M29] 1131 is turned on and global bit_n 1150 is discharged with a limited swing. When a bit to be read is logic LOW, PMOS transistor [M38] 1130 is turned on, and the global bit 1151 is discharged with a limited swing. This charge-sharing scheme can result in very little power consumption, because only the charge that causes the limited voltage swing on the global bitlines 1150, 1151 is discharged to ground. That is, there is substantially no "crowbar" current. Furthermore, this aspect of

the present invention can be useful in memories where the global bitlines are multiplexed for input and output.

Kindly delete the paragraph beginning "FIG. 14 shows" at page 37, line 30 and ending "also become inactive" at page 38, line 19, and replace it with the following paragraph:

FIG. 14 shows an embodiment of this aspect of the invention herein having four pairs of columns 1402a-d (labeled as COLUMN PAIR #1, COLUMN PAIR #2, COLUMN PAIR #3 and COLUMN PAIR #4, respectively) with one redundant pair 1404. It is desirable to implement this aspect of the present invention as pairs of lines because a significant number of RAM failures occur in pairs, whether column or row. Nevertheless, this aspect of the present invention also contemplates single line redundancy. In general, the number of fuses in fuse box 1403 used to provide redundancy can be logarithmically related to the number line pairs, e.g., column pairs: $\log_2$ (number of column pairs), where the number of column pairs includes the redundant pairs as well. Because fuses tend to be large, their number should be minimized, thus the logarithmic relation is advantageous. Fuse outputs 1405 are fed into decoder circuits 1406a-d (labeled as DECODER #1, DECODER #2, DECODER #3 and DECODER #4, respectively), e.g., one fuse output per column pair. A fuse output creates what is referred to herein as a "shift pointer". The shift pointer indicates the shift signal in the column pair to be made redundant, and subsequent column pairs can then be inactivated. It is desirable that the signals 1405 from fuse box 1410 are decoded to generate shift signal 1412a-d at each column pair. When shift signal 1412a-d for a particular column pair 1402a-d location is selected, as decoded from fuse signals 1405,

shift pointer 1412a-d is said to be pointing at this location. The shift signals for this column, and all subsequent columns to the right of the column of pair shift pointer also become inactive.

Kindly delete the paragraph beginning "This aspect of the present invention" at page 38, line 20 and ending "memory yield" at page 39, line 4, and replace it with the following paragraph:

This aspect of the present invention can be illustrated additionally in FIG. 15A and FIG. 15B, by way of the aforementioned concept of "shift pointers." In FIG. 15A, three column pairs 1501, 1502, 1503, and one redundant column pair 1504 (with no connection at the bottom) are shown. The shift procedure is conceptually indicated by way of "line diagrams". The top lines 1505-1508 of the line diagrams are representative of columns 1501-1504 within the memory core while bottom line pairs 1509-1511 are the data input/output pairs from the input/output buffers. When a shift signal, such as a signal 1405 in FIG. 14, for a particular column pair 1501-1503 is logical LOW, it is preferred that the data in 1509-1511 be connected to respective column 1501-1503 directly above it by multiplexers. FIG. 15B is illustrative of having a failed column state. When shift signal is logical HIGH, such as a signal 1405 in FIG. 14, a failed column is indicated, such as column 1552. Active columns 1550, 1551 remain unfaulted, and continue to receive their data via I/O lines 1554, 1555. However, because column 1552 has failed, data from I/O buffer 1556 can be multiplexed to the redundant column pair 1553. Diagrammatically, it appears that data in are shifted left while data out from the memory core columns are shifted right. By adjusting the location of the shift pointer, which generally is determined by the state of the fuses, the unused redundant column

pair can be shifted to coincide with a nonfunctional column, e.g., column 1552, thereby repairing the column fault and boosting the fully functional memory yield.

Kindly delete the paragraph beginning "FIG. 16 illustrates" at page 39, line 13 and ending "redundant word line" at page 39, line 29, and replace it with the following paragraph:

FIG. 16 illustrates yet another aspect of the present invention, in which selector 1600 is adapted to provide a form of redundancy. Selector 1600 can include a primary decoder circuit 1605, which may be a global word line decoder, which is coupled with a multiplexer 1610. MUX 1610 can be activated by a [redundancy circuit] selector device 1620, which may be a fuse system, programmable memory, or other circuit capable of providing an activation signal 1630 to selector 1600 via MUX 1610. Selector 1600 is suitable for implementing module-level redundancy, such as that described relative to module 1200 in FIG. 12, which may be row redundancy or column redundancy for a given implementation. In the ordinary course of operation, input word line signal 1650 is decoded in decoder circuit 1605 and, in the absence of a fault on local word line 1670, the word line signal is passed to first local line 1680. In the event a fault is detected, MUX 1610, selects second local 1660, which is preferred to be a redundant word line.

Kindly delete the paragraph beginning "FIG. 18 illustrates" at page 41, line 18 and ending "operation can commence" at page 42, line 5, and replace it with the following paragraph:

FIG. 18 illustrates one particular embodiment of this aspect of the present invention, in memory structure 1800, where both local bitline sensing and local word line decoding are used, as described above. Memory structure 1800 includes memory

module 1805 which is coupled with local word line decoder 1815 and local bit sense amplifier 1820. Within memory module 1805 are a predefined number of memory cells, for example, memory cell M1825, which is coupled with local word line decoder 1815 via local word line 1810, and local bit sense amplifier 1820 via local bitlines 1830. With typical single-port functionality, local bitlines 1830 are precharged prior to both READ and WRITE operations. During a typical READ operation, predecoder 1835 activates the appropriate global word line decoder 1840, which, in turn, activates local word line decoder 1815. Once local word line decoder 1815 determines that associated memory cell 1825 is to be evaluated, it opens memory cell 1825 for evaluation, and activates local bit sense amplifier 1820. At the end of the local sensing period, local bit sense amplifier 1820 outputs the sensed data value onto global bitlines 1845. After global sense amplifier 1850 senses the data value, the data is output to the I/O buffer 1855. If a WRITE operation is to follow the READ operation, a typical single-port device would perform another precharge operation before the WRITE operation can commence.

Kindly delete the paragraph beginning "FIG. 21 illustrates" at page 49, line 11 and ending "substantially to ground" at page 50, line 14, and replace it with the following paragraph:

FIG. 21 illustrates an embodiment of the diffusion replica delay circuit 2000 in FIG. 20. Word-line activation of a memory cell frequency is pulsed to limit the voltage swing on the high capacitance bitlines, in order to minimize power consumption, particularly in wide word length memory structures. In order to accurately control the magnitude of a bitline voltage swing, dummy bitlines can be used. It is desirable that these dummy bitlines have a capacitance which is a predefined fraction of the actual

bitline capacitance. In such a device, the capacitance ratio between dummy bitlines and real bitlines can affect the voltage swing on the real bitlines. In prior art devices using dummy bitlines, a global dummy bitline for a memory block having a global reset loop has been utilized. Such prior art schemes using global resetting tends to deliver pulse widths of a duration substantially equivalent to the delay of global word-line drivers. Such an extend pulse width allows for a bitline voltage swing which can be in excess of what actually is required to activate a sense amplifier. This is undesirable in fast memory structures, because the additional, and unnecessary, voltage swing translates into a slower structure with greater power requirements. In one aspect of the present invention, dummy bitlines are preferably partitioned such that the local bitlines generally exhibit a small capacitance and a short discharge time. Word-line pulse signals of very short duration (e.g., 500 ps or less) are desirable in order to limit the bitline voltage swing. It also may be desirable to provide local reset of split dummy bitlines to provide very short word-line pulses. Replica word-line 2110 can be used to minimize the delay between activation of memory cell 2120 and related sense amplifier 2130. Such local signaling is preferred over global signal distribution on relatively long, highly capacitive word-lines. Word-line 2140 activates dummy cell 2150 along with associated memory cell M2120, which is to be accessed. Dummy cell 2150 can be part of dummy column 2160 which may be split into small groups (for example, eight or sixteen groups). The size of each split dummy group can be changed to adjust the voltage swing on the bitline. [When a dummy bitline is completely discharged, reset signal 2170 can be locally generated which pulls word-line 2140 substantially to ground.]

Kindly delete the paragraph beginning "FIG. 22A illustrates" at page 50, line 15 and ending "speed of the bus" at page 51, line 6, and replace it with the following paragraph:

FIG. 22A illustrates controlled voltage swing data bus circuit [(CVS)] 2200 which can be useful in realizing lower power, high speed, and dense interconnection buses. [CVS] 2200 can reduce bus power consumption by imposing a limited, controlled voltage swing on bus 2215. In an essential configuration, [CVS] 2000 can include inverter 2205, pMOS pass transistor T2 2210, and one nMOS discharge transistor, such as transistor T1a 2205a. Both transistors T1a 2205a, and T2 2210 can be programmed to control the rate and extent of voltage swings on bus 2215 such that a first preselected bus operational characteristic is provided in response to input signal 2220a. Additional discharge transistors T1b 2205b and T1c 2205c can be coupled with pass transistor T2 2210, and individually programmed to respectively provide a second preselected bus operational characteristic, as well as a third preselected bus operational characteristic, responsive to respective input signals 2220b, 2220c. The preselected bus operational characteristic can be for example, the rate of discharge of the bus voltage through the respective discharge transistor T1a 2205a, T1b 2205b, and T1c 2205c, such that bus 2215 is disposed to provide encoded signals, or multilevel logic, thereon. For example, as depicted in FIG. 22A, [CVS] 2200 can provide three distinct logic levels. Additional discharge transistors, programmed to provide yet additional logic levels also may be used. Thus, it is possible for bus 2215 to replace two or more lines. Concurrently with effecting a reduction in power consumption, the limited bus voltage swing advantageously tends to increase the speed of the bus.

In the claims:

1. (Amended) [A] An address decoder for a memory cell, comprising:

a. a synchronous portion, disposed to receive and [responsive] respond to a clocked signal;

b. an asynchronous portion, coupled with a line for the memory cell [an asynchronous circuit]; and

c. a feedback-resetting portion, coupled with the synchronous portion and the asynchronous portion and interposed there between, the feedback-resetting portion substantially isolating the synchronous portion from the asynchronous portion responsive to [a predetermined] an asynchronous reset signal.

2. (Amended) The decoder of Claim 1, wherein the feedback-resetting portion substantially isolates the synchronous portion from the asynchronous portion responsive to a [predetermined] monitor signal.

6. (Amended) A decoder in a memory module having a plurality of memory cell groups, comprising:

e. a signal input;

f. a first memory output coupled with a first memory cell group;

g. a second memory output coupled with a second memory cell group; and

h. a selector coupled between the signal input, the first memory output, and the second memory output, wherein the decoder decodes the first

memory cell group, and being disposed to select and decode the second memory cell group responsive to [an] a group-select signal.

7. (Amended) The decoder of Claim 6, wherein the selector comprises a multiplexer, the multiplexer selecting to decode from one of the first memory cell group and [a memory] the second memory cell group, the multiplexer being responsive to the group-select signal.

8. (Amended) The decoder of Claim 6, wherein the decoder is a row decoder disposed in a memory module having a plurality of adjacent memory rows, and wherein a first memory row and a second memory row are adjacent memory rows in the memory module, and the group-select signal is an alternative-row-select signal.

9. (Amended) The decoder of Claim 6, wherein the decoder is a column decoder disposed in a memory module having a plurality of adjacent memory columns, and wherein a first memory column and a second memory column are adjacent memory columns in the memory module, and the group-select signal is an alternative-column-select signal.

12. (Amended) A decoder in a memory module having a plurality of memory cell groups, comprising:

h. a synchronous portion, disposed to receive and [responsive] respond to a clocked signal;

i. an asynchronous portion, coupled with [an asynchronous circuit in] a selected memory cell group;

j. a feedback-resetting portion, coupled with the synchronous portion and the asynchronous portion and interposed there between, the feedback-resetting portion substantially isolating the synchronous portion from the asynchronous portion responsive to an asynchronous reset signal;

k. a signal input;

l. a first memory output coupled with a first memory cell group;

m. a second memory output coupled with a second memory cell group; and

n. a selector coupled between the signal input, the first memory output, and the second memory output, wherein the decoder decodes the first memory cell group, and being disposed to select and decode the second memory cell group responsive to [an[a group-select signal.

13. (Amended) The decoder of Claim 12, wherein the selector comprises a multiplexer, the multiplexer selecting to decode from one of the first memory cell group and [a memory] the second memory cell group, the multiplexer being responsive to the group-select signal.

14. (Amended) The decoder of Claim 12, wherein the decoder is a row decoder disposed in a memory module having a plurality of adjacent memory rows, and wherein a first memory row and a second memory row are adjacent memory rows in the memory module, and the group-select signal is an alternative-row-select signal.